DOCUMENT RESUME

ED 262 095 TM 850 578

AUTHOR Bangert-Drowns, Robert L.
TITLE The Meta-Analytic Debate.
PUB DATE 85
NOTE 12p.; Paper presented at the Annual Meeting of the American Educational Research Association (69th, Chicago, IL, March 31-April 4, 1985).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Data; Effect Size; Error of Measurement; *Literature Reviews; *Measurement Techniques; *Meta Analysis; Research Design; *Research Methodology; Sampling
IDENTIFIERS Glass (GV)

ABSTRACT

Since meta-analysis was described in 1976 (Glass) as the application of familiar experimental methods to the integration of available research, at least five coherent approaches to meta-analysis have appeared in common use. These approaches can be divided into two broad groups. In the first group (including procedures by Robert Rosenthal, Larry Hedges, and Frank Schmidt and John Hunter), meta-analysis is used to approximate data-pooling. This type of meta-analysis attempts to answer the same questions as primary research, only larger samples are used by combining information from many studies. The alternate view (shown in the procedures of Gene Glass and James Kulik) is that meta-analysis is a form of literature review. As a literature review, meta-analysis is not meant to test a hypothesis but to summarize features and outcomes of a body of research. The differences in approaches to meta-analysis indicate that the procedure is still in a period of invention and change. It is important that editors, consumers, and critics of meta-analysis know about these differences so that they can make more informed evaluations of meta-analytic findings. (Author/PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED262095

The Meta-analytic Debate

Robert L. Bangert-Drowns
The University of Michigan

Paper presented at the annual meeting
of the American Educational Research Association
Chicago, 1985

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Bangert-Drowns, R.
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

TM 850 578

Abstract

Gene Glass first described meta-analysis in 1976 as quantitative research review. Since that time, at least five coherent approaches to meta-analysis have appeared in common use. These approaches can be divided into two broad groups. In the first group, meta-analysis is used to approximate data-pooling. This type of meta-analysis attempts to answer the same questions as primary research, only larger samples are used by combining information from many studies. The alternative view is that meta-analysis is a form of literature review. As a literature review, meta-analysis is not meant to test a hypothesis but to summarize features and outcomes of a body of research. The differences in approaches to meta-analysis indicate that the procedure is still in a period of invention and change. It is important that editors, consumers, and critics of meta-analysis know about these differences so that they can make more informed evaluations of meta-analytic findings.

The Meta-analytic Debate

In any review of educational research, it is obvious that replication is a difficult task. Repeated tests of the same hypothesis commonly produce findings that vary unpredictably. The educational researcher is left to interpret different findings from similar studies. Is there a way to provide a general measure of a treatment's effectiveness? Is there a way to explain outcome variation? Can the findings of different studies be summarized in a formal or quantitative way?

Gene Glass of the University of Colorado advocated the use of meta-analysis as an answer to these questions. Glass did not think of meta-analysis as a new technique, but as the application of familiar experimental methods to the integration of available research. Since Glass's formulation in 1976, at least five different, coherent approaches to meta-analysis have appeared in common use. It is important to distinguish these approaches so that users, critics, and consumers of meta-analysis can make more informed evaluations of these reviews.

Insert Figure 1 about here

Robert Rosenthal of Harvard University presented a simple method of quantitative review in 1976. First, studies of a particular treatment are gathered. The result of each study is changed into an effect size, the difference between experimental and control groups divided by the pooled standard deviation. The effect sizes of all the studies are averaged. Next, exact one-tailed p-values and their corresponding standard normal deviates are calculated for each study's outcome. These are used to determine the combined probability of the studies' findings.

Rosenthal's procedure, therefore, produces two statistics: an average effect size and combined probability. These provide a simple indication of the central tendency and the chance likelihood of these study outcomes. What is absent in this approach is any thorough examination of relationships between study features and study outcomes. This is a serious absence because the meta-analyst is unable to explain the variation in study findings. If the meta-analyst does not assess the influence of intervening variables, it is hard to trust the generalizability of the averaged effect.

Insert Figure 2 about here

Rosenthal, in collaboration with Donald Rubin, has recently developed more sophisticated procedures. Larry Hedges of the University of Chicago has independently developed a virtually identical approach. These meta-analysts use a test of homogeneity which compares the variance of the collected effect sizes to their sampling error. If the difference between the variance and the sampling error is statistically significant, the effect sizes are regrouped according to study features. New groups are retested until they are homogeneous and are assumed to represent a single population effect. Effect sizes are averaged within the homogeneous groups.

This meta-analytic approach permits a reviewer to examine outcome variation as well as measure its central tendency. It does have a drawback, however. The test of homogeneity may be too sensitive in some cases, showing statistical significance for unimportant differences among effect sizes. In other cases, differences among effect sizes may not be statistically significant but still reflect important relationships between study features and study outcomes.

Insert Figure 3 about here

The third meta-analytic procedure, devised by John Hunter of Michigan State University and Frank Schmidt of George Washington University, abandons the tests of homogeneity in favor of real measures of variance and sampling error. That is, outcome variance attributable to sampling error is subtracted from the overall variance of the effect sizes. If the remaining variance is large (more than 25% of the total variance), the meta-analyst must then search for intervening variables.

Using Hunter and Schmidt's approach, the meta-analyst is not confused by significance tests that may obscure noteworthy patterns of variation or be too sensitive to trivial variation. An added advantage of this approach is that the meta-analyst has greater flexibility in calculating sampling error. If a group of studies provides detailed information, a reviewer can calculate sampling error more precisely than the way that it is calculated in the standard homogeneity tests.

Insert Figure 4 about here

The final two approaches, one by Glass and the other by James Kulik of the University of Michigan, differ from the other three approaches in a fundamental way. Other meta-analysts see their work as identical to the work of the original researchers, only the meta-analyst uses larger subject pools to answer the same questions. Glass and Kulik, however, are more self-consciously literature reviewers. Glass has argued that the relationships among treatment effects, research, and publicly available reports are not simple. According to this view, meta-analysts should interpret their findings conservatively, as reflecting available literature and not necessarily real effects.

Kulik and Glass treat each effect size as a simple data point. Since the effect sizes are data points, their sampling error need not be considered and familiar statistical tests, such as t-tests or ANOVA, can be applied.

Glass and Kulik are not entirely in agreement. Kulik carefully eliminates from his reviews any studies that are judged to have serious methodological flaws. Glass is more inclusive. He is concerned that the elimination of poor quality studies introduces a reviewer's biases, narrows the available sample of research, and makes it impossible to see how certain methodological features affect study outcomes.

Secondly, Glass calculates an effect size for every dependent variable. Kulik criticizes this procedure. Any study that measures many dependent variables will arbitrarily be counted more frequently in Glass's meta-analysis. Effect sizes from the same study are not independent, which complicates the application of statistical tests and the interpretation of results. Kulik, therefore, calculates only one effect size for every study.

Finally, in Glass's meta-analyses, dependent variables that measure different kinds of results are combined. For example, to assess computer-based instruction, Glass might combine effects on achievement tests, attitudes toward computers, and attitudes toward course content. The combination of such different dependent variables is difficult to understand and such results may be too far abstracted from real practice to be useful. Therefore, Kulik analyzes different dependent variables separately.

From this brief outline of alternative forms of meta-analysis, six conclusions can be drawn.

1) Meta-analysts differ over the purpose of quantitative review. Some use the method to estimate real treatment effects, others to review a literature. Until the relationships among treatment effects, research, and available reports is better understood, the conservative interpretation, that meta-analysis is a literature review, is probably most reasonable.

2) Criteria for study inclusion should be stated with sufficient clarity to make evaluation and replication by others possible.

3) Dependent variables should be analyzed separately if they measure different kinds of results. Averaged effect sizes that reflect one type of result are more understandable and usable than effect sizes from mixed dependent variables.

4) Meta-analysts should calculate an effect size for every study, not for every dependent variable. Effect sizes will therefore be independent and no small group of studies will dominate specific or general findings.

5) If tests of homogeneity are used, they probably should be complemented with the Hunter-Schmidt procedures. In this way, the statistical tests will not mask potentially important relationships that may not reach significance.

6) Rosenthal's combined probability approach is too simple to account for outcome variation and should only be used in conjunction with more sensitive procedures.

The meta-analytic debate is not over. Differences in approach are still disputed and the specific forms of future meta-analytic procedures are impossible to predict. It is certain, however, that too much valuable information is scattered in unusable form in individual studies. Quantitative research integration is essential, therefore, to enhance the power of educational research.

References

Glass, G. V., McGaw, B., & Smith, M. L. (1981). Meta-analysis in social research. Beverly Hills, CA: Sage.

Hedges, L. V., & Olkin, I. (1982). Analyses, reanalyses, and meta-analysis. Contemporary Education Review, 1, 157-165.

Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1983). Meta-analysis: Cumulating research findings across studies. Beverly Hills, CA: Sage.

Kulik, J. A., Kulik, C.-L. C., & Cohen, P. A. (1979). A meta-analysis of outcome studies of Keller's personalized system of instruction. American Psychologist, 34, 307-318.

Rosenthal, R. (1976). Interpersonal expectancy effects: A follow-up. In R. Rosenthal, Experimental effects in behavioral research (pp. 440-471). New York: Irvington.

Rosenthal, R. (1978). Combining results of independent studies. Psychological Bulletin, 85, 185-193.

Figure 1

Method suggested by Rosenthal

> Average effect sizes from all studies

> Calculate combined probability

Disadvantage

> Does not explain variation in study findings

Figure 2

Method suggested by Hedges and by Rosenthal and Rubin

> Apply tests of homogeneity to effect sizes

> Study relationships between study features and effect sizes

Disadvantage

> Significance tests may mask important relationships

Figure 3

Method suggested by Hunter and Schmidt

> Remove outcome variance due to sampling error

> Search for intervening variables

Advantages

> Avoids significance testing

> Allows flexibility in calculating sampling error

Figure 4

Methods suggested by Glass and Kulik

Agree that:

> Meta-analysis is a literature review

> Familiar parametric tests are applied to effect sizes

Disagree about:

> Study inclusion decisions

> Calculation of effect size for every dependent variable

> Combining different kinds of dependent variables